

# **ENHANCED CLUSTERING ALGORITHMS FOR GRAY-SCALE IMAGE SEGMENTATION**

**FASAHAT ULLAH SIDDIQUI**

**UNIVERSITI SAINS MALAYSIA**

**2012**

**ENHANCED CLUSTERING ALGORITHMS FOR GRAY-SCALE IMAGE  
SEGMENTATION**

**by**

**FASAHAT ULLAH SIDDIQUI**

**Thesis submitted in fulfillment of the  
requirements for the degree of Master of  
Science**

**APRIL 2012**

## **ACKNOWLEDGEMENT**

All thanks are to Allah alone for bestowing me with intellect and patience which has made all of this possible.

I would like to express my deepest appreciation to my supervisor, Assoc. Prof. Dr. Nor Ashidi bin Mat Isa for his invaluable guidance and encouragement throughout the completion of the research project. His contributions and detailed comments have been of great value to me. Furthermore, I would like to thank him for improving and polishing my technical writing skills.

I would also like to take this opportunity to thank my parents and grandparents for all the sacrifices that they have endured and for their unconditional love and support. For all of this and much more, I dedicate this thesis to them. My thanks are also to my siblings for their encouragement and support.

Next to this, I would like to convey my humble gratitude to Dr. Abid Yahya and others friends for their encouragement and faithful advices. I am also thankful to my fellow friends of the Imaging and Intelligent System Research Team (ISRT) for their support.

Last but certainly not the least, I would like to acknowledge, this work has been partially supported by short term grant, Universiti Sains Malaysia entitled “Fuzzy Logic Based Segmentation Techniques for Determination of Breast Tumor Region on Mammogram Image”, and Research University (Individual) Grant, Universiti Sains Malaysia titled “Study on Capability of FTIR Spectral Characteristics for Development of Intelligent Cervical Precancerous Diagnostic System”.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENT</b> .....	ii
<b>TABLE OF CONTENTS</b> .....	iii
<b>LIST OF TABLES</b> .....	vi
<b>LIST OF FIGURES</b> .....	ix
<b>LIST OF ABBREVIATIONS</b> .....	xiv
<b>ABSTRAK</b> .....	xv
<b>ABSTRACT</b> .....	xvii

### CHAPTER 1 - INTRODUCTION

1.1 Background .....	1
1.2 Clustering Technique .....	3
1.3 Problem Statement .....	6
1.4 Thesis Objectives .....	7
1.5 Thesis Outline .....	8

### CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction .....	10
2.2 Image Segmentation .....	11
2.2.1 Basic Concept .....	11
2.2.2 Image Segmentation Techniques .....	13
2.2.2(a) Thresholding Technique .....	13
2.2.2(b) Region Growing Technique .....	16
2.2.2(c) Split-and-Merge Technique .....	18
2.2.2(d) Clustering Technique .....	20
2.3 Clustering-based Segmentation Technique .....	23
2.3.1 Hard Clustering Technique .....	23

2.3.2	Soft Clustering Technique.....	26
2.3.3	Application of Clustering Technique .....	29
2.4	Implementation and Validation.....	32
2.4.1	Analysis Methods .....	34
2.4.1(a)	Qualitative Analysis .....	34
2.4.1(b)	Quantitative Analysis .....	35
2.4.2	Selected Clustering Algorithms.....	38
2.4.2(a)	K-Means Clustering Algorithm.....	39
2.4.2(b)	Moving K-Means Clustering Algorithm .....	43
2.4.2(c)	Adaptive Moving K-Means Algorithm .....	49
2.4.2(d)	Fuzzy C-Means Clustering Algorithm .....	53
2.4.2(e)	Adaptive Fuzzy Moving K-Means Clustering Algorithm.....	57
2.4.2(f)	Adaptive Fuzzy K-Means Clustering Algorithm .....	61
2.5	Summary.....	65
<b>CHAPTER 3 - METHODOLOGY</b>		
3.1	Introduction.....	67
3.2	Optimized K-Means Algorithm .....	68
3.3	Enhanced Moving K-Means Algorithm.....	74
3.4	Outlier Rejection Fuzzy C-Means Algorithm.....	82
3.5	Data Sample and Analysis .....	85
3.6	Summary .....	89
<b>CHAPTER 4 - EXPERIMENTAL RESULTS AND DISCUSSIONS</b>		
4.1	Introduction.....	91
4.2	Testing the Capability of the Proposed Clustering Algorithms .....	92
4.2.1	Optimized K-Means Clustering Algorithm.....	92
4.2.2	Enhanced Moving K-Means Clustering Algorithm .....	95

4.2.3	Outlier Rejection Fuzzy C-Means Clustering Algorithm.....	97
4.3	Performance Comparison.....	100
4.3.1	Qualitative Analysis .....	101
4.3.2	Quantitative Study .....	140
4.3.2(a)	Number of Dead Centroids.....	141
4.3.2(b)	Processing Time .....	144
4.3.2(c)	MSE Experimental Analysis .....	148
4.3.2(d)	INTER Experimental Analysis.....	151
4.3.2(e)	VXB Experimental Analysis .....	155
4.3.2(f)	$F(I)$ Experimental Analysis .....	158
4.3.2(g)	$F'(I)$ Experimental Analysis.....	161
4.3.2(h)	$Q(I)$ Experimental Analysis.....	165
4.4	Summary.....	169
<b>CHAPTER 5 - CONCLUSION AND FUTURE WORKS</b>		
5.1	Conclusion .....	170
5.2	Future Works .....	172
<b>REFERENCES.....</b>		174
<b>LIST OF PUBLICATIONS.....</b>		188

## LIST OF TABLES

Table 2.1 Characteristics comparison between segmentation techniques. ....	22
Table 2.2 Characteristics of the several review clustering algorithms.....	33
Table 2.3 Results of quantitative analysis for <i>Bridge</i> image after applying with the KM algorithm. ....	43
Table 2.4 Results of quantitative analysis for <i>Harbor House</i> image after applying the MKM algorithm.....	49
Table 2.5 Results of quantitative analysis for <i>Gantry Crane</i> image after applying the AMKM algorithm.....	52
Table 2.6 Results of quantitative analysis for <i>Football</i> image after applying the FCM algorithm.....	56
Table 2.7 Results of quantitative analysis for <i>Light House</i> image after applying the AFMKM algorithm.....	61
Table 2.8 Results of quantitative analysis for <i>Monument Building</i> image after applying the AFKM algorithm. ....	64
Table 4.1 Results of quantitative analysis for <i>Bridge</i> and <i>Light House</i> images after applying the OKM algorithm. ....	94
Table 4.2 Results of quantitative analysis for <i>Gantry Crane</i> image after applying the EMKM-1 and EMKM-2 algorithms.....	97
Table 4.3 Results of quantitative analysis for <i>Football</i> image after applying the ORFCM algorithm.....	100
Table 4.4 Number of dead centroids produced using different clustering algorithms for selected standard images.....	142
Table 4.5 Number of dead centroids produced using different clustering algorithms for selected cervical cells and malaria parasite images. ....	143
Table 4.6 Total numbers of dead centroids produced using different clustering algorithms for 103 standard, 150 cervical cells, and 150 malaria parasite images.....	144
Table 4.7 Processing Time (in seconds) for selected standard images after applying different clustering algorithms.....	145
Table 4.8 Processing time (in seconds) for selected cervical cells and malaria parasite images after applying different clustering algorithms. ....	146

Table 4.9 Average processing time (in seconds) for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying different clustering algorithms. ....	147
Table 4.10 MSE value for selected standard images after applying different clustering algorithms. ....	149
Table 4.11 MSE value for selected cervical cells and malaria parasite images after applying different clustering algorithms. ....	150
Table 4.12 Average MSE value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying different clustering algorithms. ....	151
Table 4.13 INTER value for selected standard images after applying the different clustering algorithms. ....	153
Table 4.14 INTER value for selected cervical cells and malaria parasite images after applying the different clustering algorithms. ....	154
Table 4.15 Average INTER value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying the different clustering algorithms. ....	155
Table 4.16 VXB value for selected standard images after applying the different fuzzy-based clustering algorithms. ....	156
Table 4.17 VXB value for selected cervical cells and malaria parasite images after applying the different fuzzy-based clustering algorithms. ....	156
Table 4.18 Average VXB value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying the different fuzzy-based clustering algorithms. ....	157
Table 4.19 $F(I)$ value for selected standard images (in $1 \times 10^7$ ) after applying the different clustering algorithms. ....	159
Table 4.20 $F(I)$ value for selected cervical and malaria parasite images (in $1 \times 10^4$ ) after applying the different clustering algorithms. ....	160
Table 4.21 Average $F(I)$ value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying the different clustering algorithms. ....	161
Table 4.22 $F'(I)$ value for selected standard cell images after applying the different clustering algorithms. ....	162
Table 4.23 $F'(I)$ value for selected cervical cells and malaria parasite images after applying the different clustering algorithms. ....	163



Table 4.24 Average $F'(I)$ value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying the different clustering algorithms. ....	164
Table 4.25 $Q(I)$ value for selected standard images after applying the different clustering algorithms. ....	166
Table 4.26 $Q(I)$ value for selected cervical cells and malaria parasite images after applying the different clustering algorithms. ....	167
Table 4.27 Average $Q(I)$ value for 103 standard, 150 cervical cells, and 150 malaria parasite images after applying the different clustering algorithms. ....	168

## LIST OF FIGURES

Figure 1.1 Block diagram of the image processing system. ....	2
Figure 1.2 Vector spaces for (a) color image and (b) gray-scale image. ....	2
Figure 1.3 Flow chart of the clustering process. ....	4
Figure 1.4 Data group into their respective cluster by (a) hard clustering algorithm and (b) soft clustering algorithm. ....	5
Figure 2.1 Segmentation process for the (a) color, (b) gray intensity and (c) texture images. ....	12
Figure 2.2 General classifications of the image segmentation technique. ....	13
Figure 2.3 Implementation of the thresholding technique for a various threshold values (a) original image, (b) histogram of an image, (c) thresholded at 111, (d) thresholded at 157, and (e) thresholded at 226. ....	15
Figure 2.4 Implementation of the thresholding technique with multiple threshold values (a) original image and (b) segmented image. ....	15
Figure 2.5 Implementation of region growing technique (a) location of initial seed pixel, (b), (c), and (d) the growing process of seed pixel to form a region. .....	17
Figure 2.6 The example of region growing application (a) original image (b) segmented image. ....	17
Figure 2.7 Implementation of the split-and-merge technique. ....	19
Figure 2.8 Illustration of basic clustering process; (a) initial condition when iteration=0, (b) condition of clustering process after iteration $t$ , and (c) final condition after clustering process is completed. ....	21
Figure 2.9 Result of segmentation process using FCM clustering technique; (a) original image (b) resultant image. ....	21
Figure 2.10 General classification of clustering technique. ....	23
Figure 2.11 Flow chart for the implementation of KM algorithm. ....	40
Figure 2.12 The data point with equal distance to its adjacent clusters. ....	41
Figure 2.13 Implementation of the KM algorithm for different number of cluster ' $k$ '; (a) original <i>Bridge</i> image, (b) histogram of original image, (c) resultant	

image at $k=3$ , (d) histogram of the resultant image, (e) resultant image at $k=6$ , and (f) histogram of the resultant image. ....	42
Figure 2.14 Flow chart for the implementation of MKM algorithm.....	46
Figure 2.15 Implementation of the MKM algorithm; (a) original <i>Harbor House</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image.....	48
Figure 2.16 Flow chart for the implementation of AMKM algorithm.....	50
Figure 2.17 Implementation of the AMKM algorithm; (a) original <i>Gantry Crane</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image.....	52
Figure 2.18 Flow chart for the implementation of FCM algorithm. ....	54
Figure 2.19 Membership function generated by the FCM algorithm for data of intensity range 1 to 120. ....	55
Figure 2.20 Implementation of the FCM algorithm; (a) original <i>Football</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image. ....	57
Figure 2.21 Flow chart for the implementation of AFMKM algorithm. ....	59
Figure 2.22 Implementation of the AFMKM algorithm; (a) original <i>Light House</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image.....	60
Figure 2.23 Flow chart for the implementation of AFKM algorithm. ....	63
Figure 2.24 Implementation of the AFKM algorithm; (a) original <i>Monument Building</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image. ....	65
Figure 3.1 Flow chart for the implementation of proposed OKM algorithm.....	73
Figure 3.2 Flow chart for the implementation of proposed EMKM-1 algorithm. ....	80
Figure 3.3 Flow chart for the implementation of proposed EMKM-2 algorithm. ....	81
Figure 3.4 Flow chart for the implementation of proposed ORFCM algorithm.....	84
Figure 3.5 Example of abnormal and normal cervical cells; (a) HSIL, (b) LSIL, and (c) normal squamous cell.....	87
Figure 3.6 Illustration of the segmented cervical cell images taken from Figure 3.5 (c); cytoplasm in red and nucleus in blue colors. ....	87

Figure 3.7 Example of three types of malaria parasites; (a) plasmodium falciparum (PF), (b) plasmodium vivax (PV), and (c) plasmodium malariae (PM)....	87
Figure 3.8 Illustration of the segmented malaria parasites images taken from Figure 3.7 (a); RBC in red and malaria parasites in blue colors. ....	87
Figure 4.1 Implementation of the OKM algorithm; (a) original <i>Bridge</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image.....	93
Figure 4.2 Implementation of the OKM algorithm; (a) original <i>Light House</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image.....	94
Figure 4.3 Implementation of EMKM-1 and EMKM-2 algorithms; (a) original <i>Gantry Crane</i> image, (b) histogram of original image, (c) EMKM-1 resultant image, (d) histogram of EMKM-1 resultant image, (e) EMKM-2 resultant image, and (f) histogram of EMKM-2 resultant image. ....	96
Figure 4.4 Membership function generated by the ORFCM algorithm for data of intensity range (a) 1 to 120 and (b) 1 to 60. ....	98
Figure 4.5 Implementation of the ORFCM clustering algorithm; (a) original <i>Football</i> image, (b) histogram of original image, (c) resultant image, and (d) histogram of resultant image. ....	99
Figure 4.6 Segmentation results in 3 clusters for the image <i>Lake</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	103
Figure 4.7 Segmentation results in 3 clusters for the image <i>Building</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	105
Figure 4.8 Segmentation results in 4 clusters for the image <i>Aircraft</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	108
Figure 4.9 Segmentation results in 4 clusters for the image <i>Harbor House</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	110

Figure 4.10 Segmentation results in 5 clusters for the image <i>House</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	114
Figure 4.11 Segmentation results in 5 clusters for the image <i>Camera Man</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	116
Figure 4.12 Segmentation results in 6 clusters for the image <i>Fruit Table</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	118
Figure 4.13 Segmentation results in 6 clusters for the image <i>Alpine Skiing</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM. ....	120
Figure 4.14 Segmentation results in 3 clusters for the image <i>HSIL-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	122
Figure 4.15 Segmentation results in 3 clusters for the image <i>HSIL-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	124
Figure 4.16 Segmentation results in 3 clusters for the image <i>LSIL-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	125
Figure 4.17 Segmentation results in 3 clusters for the image <i>LSIL-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	127
Figure 4.18 Segmentation results in 3 clusters for the image <i>N-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	129

Figure 4.19 Segmentation results in 3 clusters for the image <i>N-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	130
Figure 4.20 Segmentation results in 3 clusters for the image <i>PF-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	133
Figure 4.21 Segmentation results in 3 clusters for the image <i>PF-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	134
Figure 4.22 Segmentation results in 3 clusters for the image <i>PV-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	136
Figure 4.23 Segmentation results in 3 clusters for the image <i>PV-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	137
Figure 4.24 Segmentation results in 3 clusters for the image <i>PM-1</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	139
Figure 4.25 Segmentation results in 3 clusters for the image <i>PM-2</i> after applying different clustering algorithms; (a) Original image, (b) KM, (c) FCM, (d) MKM, (e) AMKM, (f) AFMKM, (g) AFKM, (h) OKM, (i) EMKM-1, (j) EMKM-2, and (k) ORFCM.....	140

## LIST OF ABBREVIATIONS

AFKM	Adaptive Fuzzy K-Means
AFMKM	Adaptive Fuzzy Moving K-Means
AMKM	Adaptive Moving K-Means
Bootstrap WFCM	Bootstrap Weighted Fuzzy C-mean
EMKM-1	Enhanced Moving K-Means-1
EMKM-2	Enhanced Moving K-Means-2
FCM	Fuzzy C-Means
HSIL	High-grade squamous intraepithelial lesion
KM	K-Means
Lp Norm FCM	Lp Norm Fuzzy C-Means
LSIL	Low-grade squamous intraepithelial lesion
MKM	Moving K-Means
MSE	Mean Square Error
OKM	Optimized K-Means
ORFCM	Outlier Rejection Fuzzy C-Means
PAM	Partitioning Around Medoids
PF	Plasmodium Falciparum
PK	Plasmodium Knowlesi
PM	Plasmodium Malaria
PO	Plasmodium Ovale
PV	Plasmodium Vivax
RCFCM	Revival Check Fuzzy C-Means
ROI	Region of Interest
S-FCM	Suppress-Fuzzy C-Means
VXB	Validity Xie-Beni
WFCM	Weighted Fuzzy C-mean

# **PENAMBAHBAIKAN ALGORITMA PENGELOMPOKAN UNTUK PERUASAN IMEJ SKALA KELABU**

## **ABSTRAK**

Algoritma pengelompokan digunakan secara meluas sebagai kaedah tanpa seliaan untuk peruasan imej dalam diagnosis perubatan, pengimejan satelit dan sistem biometrik. Algoritma ini dipilih kerana ia mudah dilaksanakan, memerlukan masa pengkomputeran yang sedikit dan kurang peka terhadap hingar dan artifak. Walau bagaimanapun, dalam sesetengah kes, penggunaan algoritma pengelompokan konvensional akan mengakibatkan masalah terlebih-ruas dan tidak dapat memelihara kawasan pilihan (iaitu objek). Masalah pertama timbul kerana algoritma peruasan tertentu adalah peka terhadap unsur luaran, manakala masalah kedua adalah disebabkan oleh fenomena pusat mati dan pusat terperangkap pada kawasan tidak aktif. Oleh itu, empat algoritma pengelompokan baru iaitu Purata-K Teroptima (PKT), Purata-K Boleh Gerak Dipertingkatkan-1 (PKBGD-1), Purata-K Boleh Gerak Dipertingkatkan-2 (PKBGD-2) dan Purata-C Kabur Penolakan Unsur Luaran (PCKPUL) dicadangkan. Algoritma PKT mempunyai keupayaan untuk membezakan di antara kelompok tanpa ahli dan kelompok varians sifar. Umpukan data kepada kelompok-kelompok ini diubahsuai daripada algoritma Purata-K. Algoritma PKBGD-1 dan PKBGD-2 mengubahsuai konsep pemindahan data yang digunakan dalam Purata-K Boleh Gerak Adaptif (PKBGA) agar mampu mengurangkan masalah-masalah yang dinyatakan. Selain itu, algoritma PCKPUL pula mengurangkan operator eksponen boleh suai jarak Euclidean dalam fungsi keahlian kabur untuk meningkatkan keupayaan Purata-C Kabur dengan meminimumkan kawasan bertindih yang bertujuan mengurangkan kesan unsur luaran. Prestasi



algoritma pengelompokan yang dicadangkan ini dibandingkan dari segi kualiti dan kuantiti dengan beberapa algoritma pengelompokan terkini. Keputusan yang diperolehi membuktikan bahawa algoritma pengelompokan yang dicadangkan ini berupaya mengatasi algoritma pengelompokan konvensional dengan menghasilkan lebih banyak kawasan homogen pada imej yang diruas dengan bentuk objek yang lebih terpelihara dan pemuliharaan sisi yang lebih jelas. Secara keseluruhan, algoritma PKT menunjukkan hasil peruasan terbaik berbanding algoritma-algoritma lain yang diuji.

# **ENHANCED CLUSTERING ALGORITHMS FOR GRAY-SCALE IMAGE SEGMENTATION**

## **ABSTRACT**

The clustering algorithms are widely used as an unsupervised method for image segmentation in medical diagnosis, satellite imaging and biometric systems. The algorithms are chosen since they are easy to be implemented, required low computational time and less sensitive to noise and artifacts. However, in some cases the conventional clustering algorithms introduce over-segmentation problems and unable to preserve the region of interest (i.e. objects). The first problem occurs because of certain clustering algorithms are sensitive to outliers, meanwhile the latter problem is due to the dead centroids and trapped centroids at non-active regions phenomenon. Thus, four new clustering algorithms namely the Optimized K-Means (OKM), Enhanced Moving K-Means-1(EMKM-1), Enhanced Moving K-Means-2(EMKM-2) and Outlier Rejection Fuzzy C-Means (ORFCM) are proposed. The OKM algorithm has a capability to differentiate between the empty and the zero variance clusters. The assignment of the data to these clusters are redesigned from that of the conventional K-Means clustering algorithm. On the other hand, the EMKM-1 and EMKM-2 algorithms reformed the data transferring concept of the conventional Adaptive Moving K-Means (AMKM) which is able to avoid the aforementioned problems. Furthermore, the ORFCM algorithm employed the adaptable exponent operator of Euclidean distance in fuzzy membership function in order to improve the capability of the Fuzzy C-Means (FCM) by minimizing the overlapping regions to moderate the outlier effects. The performance of the proposed clustering algorithms is compared qualitatively and quantitatively with several state-

of-the-art clustering algorithms. The results conclude that the proposed clustering algorithms outperform the conventional clustering algorithms by producing more homogenous regions in an image with better in shape and sharp edges preservation. In addition, the OKM algorithm attains the best results among the other tested algorithms.

# CHAPTER 1

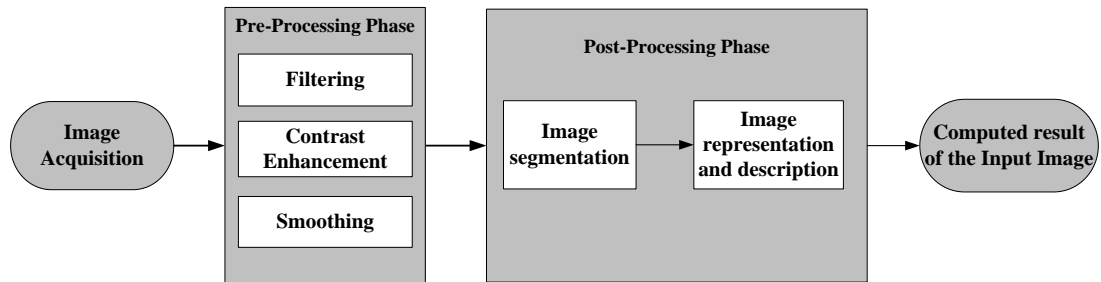
## INTRODUCTION

### 1.1 Background

In general, segmentation is defined as partitioning a given image into several homogenous regions based on certain attributes such as, texture, gray-scale and color intensity (Khotanzad and Bouarfa, 1990). The concept of image segmentation could be found in many applications such as face recognition (Garcia and Tziritas, 1999, Ambalakat, 2005), iris or fingerprint recognition (Lee et al, 2007, Tae-Hong and Rae-Hong, 2008, Balti et al, 2011), cells or tissue volume and area calculation (Brown et al., 1997, Naik et al., 2008), and satellite images application (Carleer et al., 2005).

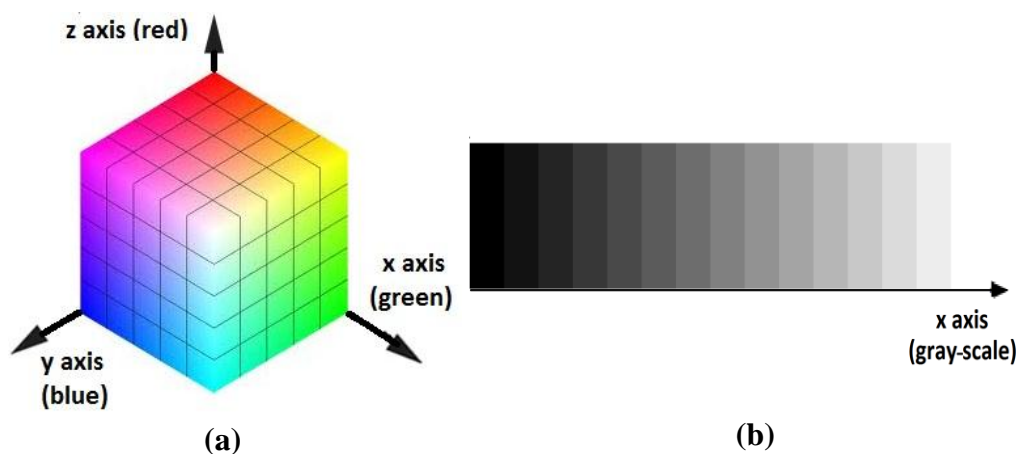
Figure 1.1 illustrates the basic processing steps that are required for image understanding in image processing system. In pre-processing phase, several tasks will be performed according to the computer vision applications requirements. Usually, the image enhancement, noise filtering and image smoothing are implemented to ease the vision task and produce the visually acceptable image. In post processing phase, the image segmentation plays a fundamental role as the initial step before applying images to higher-level operations, i.e. image representation and description (Gonzalez et al., 2004). Therefore, the accuracy of image processing system highly depends on the performance of the image segmentation step.

Generally, it classifies the data into several subgroups. After the image segmentation, the data representation and description task is performed. It converts the data in suitable form such that it becomes ease to be understood by computer or machine.



**Figure 1.1** Block diagram of the image processing system.

There are two forms of images namely the gray-scale and color images which are employed for the image segmentation process. Normally, the color image is represented by green, blue and red intensity level of a visible spectrum. The term gray-scale is generally used to describe monochromatic intensity because it ranges from black to gray and finally to white (Gonzalez et al., 2004). Figure 1.2 illustrates the number of vector spaces that have been occupied by the gray-scale and color images. Each vector space contains 256 shades with size of 8 bits.



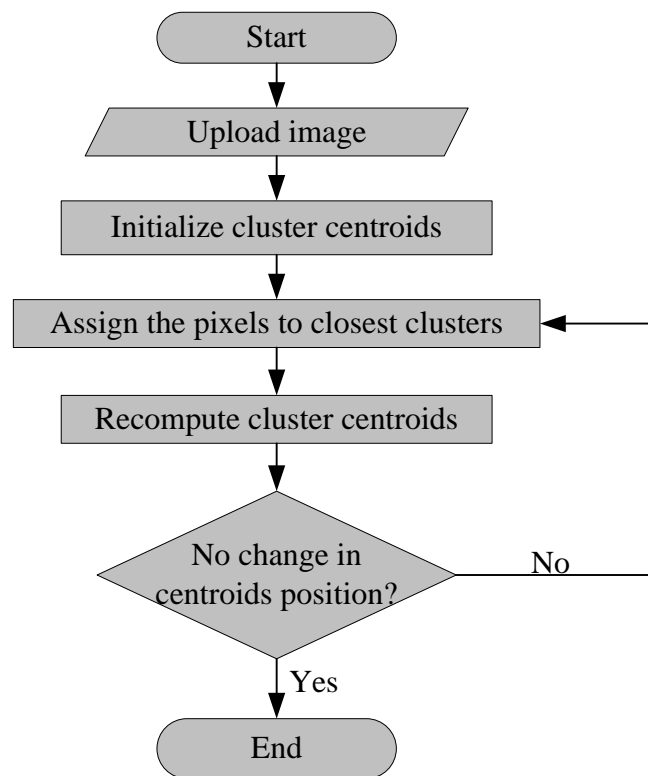
**Figure 1.2** Vector spaces for (a) color image and (b) gray-scale image.

The use of 24 bit color images for image processing is preferable as compared to the 8 bit gray-scale image. One of the main reasons is that the color image can provide additional information (Chen and Lu, 2002, Chi and Wang, 2000, Tse-Wei et al., 2008). Typically, the application area is limited for the color images because of system specifications such as limited capacity for data acquisition or storage, which prohibits large processing time (Gasparri et al., 2011). For example, in airborne and satellite imaging system, the segmentation process is limited mainly by the computational time (Guarnieri and Vettore, 2002). For that reason, the gray-scale images have been used instead of color images for satellite image segmentation (Das and Konar, 2009, Lucieer and Stein, 2002, Zhu et al., 2005, Das et al., 2006). In addition, the color image is more sensitive to noise as noise can affect three vector spaces as compared to one vector space in gray-scale image (Gevers and Smeulders, 2000). Furthermore, in medical application most of raw data are obtained in form of gray-scale images, e.g. ultrasound images, X-rays images, computed tomography (CT) images, magnetic resonance imaging (MRI) images, electron microscopic images etc. Thus, the gray-scale image segmentation in medical application is still an active research area, where a lot of research studies have been carried out (Bonnet et al., 2002, Chang et al., 2005, Ghosh and Mitchell, 2006, Pohle and Toennies, 2001, Segyeong et al., 2004, Bengtsson et al., 2004).

## **1.2 Clustering Technique**

From machine learning perspective, clustering is an unsupervised technique which classifies the pattern without using training of data. In image segmentation, clustering is used to examine the similarities and dissimilarities of objects in an

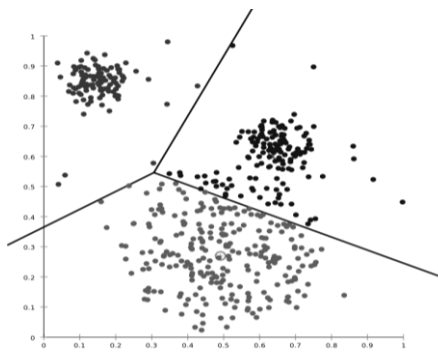
image. The pixels are naturally grouped into cluster, where the characteristics of objects are matched. In general, clustering is a centroid-based algorithm which begins with a guess about the final solution and then updates the centroids position by iteratively computing the mean of the pixels belonging to the clusters. The process will be stopped when the solution reaches a global optimum location. Figure 1.3 illustrates the basic concept of the clustering process.



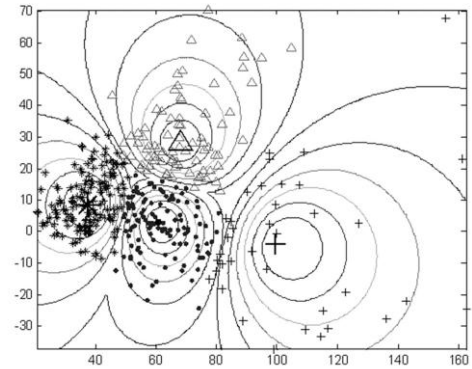
**Figure 1.3** Flow chart of the clustering process.

The clustering technique is mainly classified into two categories namely hard clustering technique and soft clustering technique. The unsupervised clustering techniques use the degree of pixels belongingness to certain clusters, which is iteratively updated in order to train themselves. In hard clustering technique, the K-

Means (KM) is the earliest existing algorithm that employed the hard membership function (MacQueen, 1967). The hard membership function is calculated according to the Euclidean distance between the pixels and user-defined number of clusters. The membership degree value is one for a pixel in a specific cluster that has the shortest Euclidean distance and zero in the other clusters. Thus, the pixel will be bounded to become a member of only one cluster as shown in Figure 1.4 (a). Finally, the new centroid value is calculated by the mean of the pixels in a cluster. This process is continued until there is no further significant change. The Fuzzy C-Means (FCM) generalizes the hard membership function and allows the pixels to be segmented according to fuzzy concept (Bezdek, 1981). In fuzzy concept, the soft membership is employed, which allows the pixel to partially allocate to all clusters instead of single cluster as shown in Figure 1.4 (b). The centroid is measured by the mean value of all pixels, weighted by their degree of membership to each cluster (Hamerly and Elkan, 2002).



(a)



(b)

**Figure 1.4** Data group into their respective cluster by (a) hard clustering algorithm and (b) soft clustering algorithm.



### 1.3 Problem Statement

The objective of hard and soft clustering algorithms is that the final solution will converge at a global optimum location or as near as possible to it. The K-means is an earliest unsupervised hard clustering algorithm and is still the most popular method due to its low execution time and ease in implementation (Hamerly and Elkan, 2002). Sometimes, in the process of searching for the global optimum location, centroids often get trapped at local minima (Mashor, 2000). Moreover, in worst case, the dead centroid is produced due to trapped centroid is not updated in next iteration. Unfortunately, this problem could arise if the centroid value has been initialized too far from the active location or region (Mashor, 2000, Mat Isa et al., 2010).

One of the possible solutions for the aforementioned problem is to keep the centroids in the active region by transferring the certain members of the active clusters to cluster that has been trapped at non-active location (Mashor, 2000). The MKM and AMKM employ transferring process that imposes the vital impact on image segmentation. By using this concept, the dead centroid problem is purged but due to improper pixels transferring the trapped centroid problem is still unsolved. Therefore, the forceful transferring of elements from one cluster to the other must be accurate otherwise it will lead to improper segmentation result. Furthermore, in the algorithm with hard membership function, the pixel that has the equal distance to two or more adjacent clusters could be assigned to the higher variance cluster and the lower variance cluster will not be trained or updated in the learning process. Due to these problems, the algorithms could fail to homogenously segment an image.

On the other hand, the soft membership makes the FCM less sensitive to initialization and could avoid dead centroid problem of the KM. However, in some cases, the FCM could not minimize the intra-cluster variance and could not maximize the inter-cluster variance due to the overlapping of regions or sensitivity to the outliers. This is due to the Euclidean distance concept which employed during calculation of the membership value is very sensitive to outlier. Thus, as the membership has great impact on the homogenous image segmentation, the outlier sensitivity of fuzzy membership must be reduced by decreasing its value for the far located pixels.

#### **1.4 Thesis Objectives**

Based on the aforementioned problem statement in Section 1.3, this study focuses on the following objectives:

1. To modify the hard membership concept as employed by the conventional KM algorithm in order to remove the dead centroid problem and classify an image into more homogenous regions.
2. To enhance the moving or transferring concept of the conventional MKM and AMKM algorithms (i.e. certain pixels or members of the cluster with high cluster variance value are forced to become the members of the clusters with the smallest cluster variance value and the nearest cluster,

respectively) such that both dead centroid and centroid trapped problems will be reduced.

3. To modify the membership function of the Fuzzy C-Means clustering algorithms in order to reduce the effect of outliers during the segmentation process.

## **1.5 Thesis Outline**

The work described in this thesis is organized into 5 chapters. Chapter 1 gives an overview on the image processing systems and specifically discusses its image segmentation step. In addition, the pros and cons of the vector space for two types of images on image segmentation are outlined. Finally, the problems associated with the conventional hard and soft unsupervised clustering algorithms and the objectives of this study are also presented.

Chapter 2 reviews the current hard and soft unsupervised clustering algorithms for image segmentation. The relevant previous work in both categories for reducing the outlier effect, avoid the centroid being trapped at non-active location, dead centroid and improving the computational speed are reviewed. The step of the implementation for the selected latest clustering algorithms with their limitations will be discussed in detail. In addition, the image segmentation evaluation methods are also presented in this chapter.

Chapter 3 describes the four proposed unsupervised clustering algorithms that are capable of reducing the primary problems of conventional algorithms (sensitive to outlier, dead centroid and trapped centroid at local minima problems). These improvements have significant impact in reducing the fundamental issues (i.e. segments an image into homogenous region with clear and sharp edges) of image segmentation. In addition, the detail steps of the implementation for the proposed clustering algorithms are presented.

Chapter 4 explains the capability of the proposed clustering by improving the previous existing problems of the conventional clustering algorithms. Finally, the obtained results after applying the proposed and comparative clustering algorithms on 103 standard images with 3, 4, 5, and 6 clusters and 300 microscopic images with 3 clusters are compared qualitatively and quantitatively. This comparison has been performed to evaluate the capability of the proposed clustering algorithms and to verify if they could segment the tested images homogeneously with perfect shape and sharp edges.

Chapter 5 concludes the proposed work in this thesis and presents the scope for future works.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The success of advance image processing modules such as object representation and description is highly influenced by the accuracy of image segmentation process (Gonzalez et al., 2004). In general, clustering technique is most commonly used for image segmentation. Several studies have been carried out to strengthen the performance of clustering algorithms (Pedrycz, 1996, Mashor, 2000, Kanungo et al., 2002, Likas et al., 2003, Kokkinos and Maragos, 2008). However, these clustering algorithms have been reported to produce inadequate segmentation results. Moreover, the previously proposed enhanced versions of the Fuzzy C-Means (FCM) algorithm generated significant segmentation for application with more than one feature vectors (Wang et al., 2004, Hung et al., 2008). For the large feature vector, these algorithms required high cost of function or long execution time. Similar to other clustering algorithms, these algorithms are also sensitive to initial value of parameters (i.e. initial centroid value of clusters) (Dehariya et al., 2010).

In addition, several improvements have also been proposed to the conventional K-Means clustering algorithm in order to significantly reduce the empty cluster (or known as dead centroid) problem by modifying the members assigning process (Mashor, 2000, Zalik, 2008). However, insignificant convergence problem (i.e. final

solution may converge to poor local minima) and centroids being trapped at non-active region problem are not completely avoided by those K-Means based clustering algorithms (Mat Isa et al., 2010).

Chapter 2 specifically deals with these issues, and presents a comprehensive review on clustering algorithms proposed by other researchers. The rest of Chapter 2 is as follows: Section 2.2 highlights the basic concept of image segmentation techniques and its applications. The well-known clustering algorithms are discussed in Section 2.3. The applications of clustering algorithms for image segmentation are then presented. The implementation of the selected clustering algorithms is presented in Section 2.4. In addition, the limitations of those clustering algorithms are also explained in that section. Finally, Section 2.5 presents the summary.

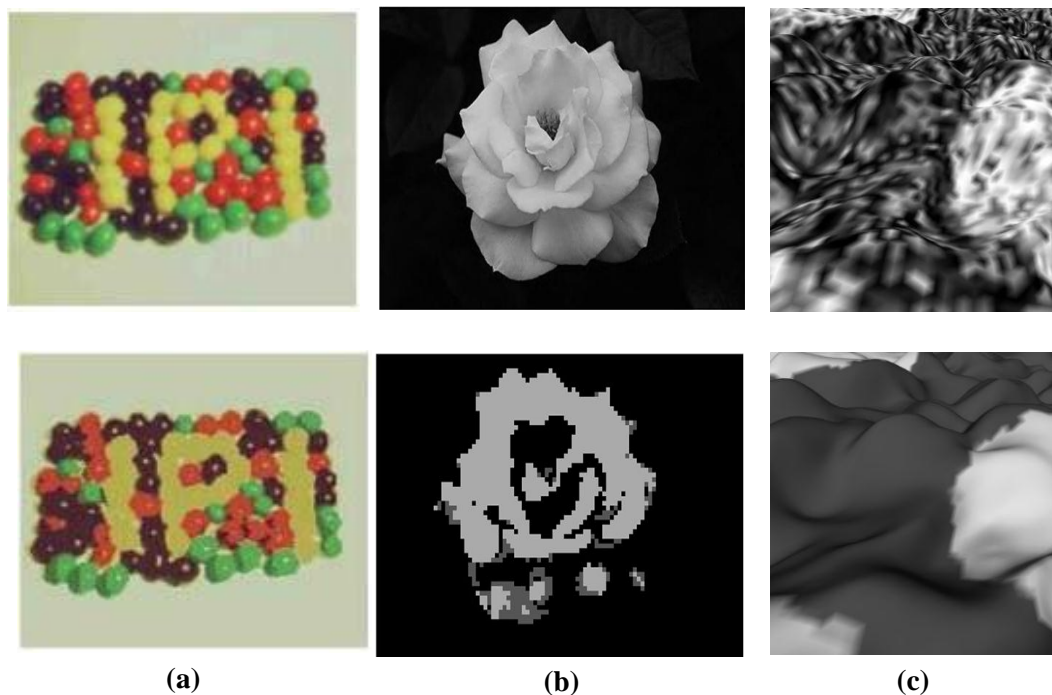
## **2.2 Image Segmentation**

### **2.2.1 Basic Concept**

Image segmentation is a process of subdividing a color, gray or texture image into multiple homogenous regions of interest, on the basis of similarity criteria as shown in Figure 2.1. It can also be defined as partitioning of an image into non-overlap regions, followed by assigning a label (i.e. intensity) to each region. The level to which subdivision or partitioning is carried out depends on the problem being solved (Kaur and Kamra, 2011). Generally, the segmentation process should stop when the regions of interest (ROI) have been isolated. As an initial step, the image

segmentation can affect the success or failure of post image processing phase such as data representation, data description etc. Please refer to Figure 1.1 for an illustration. Hence, considerable attention should be given in order to improve the performance of segmentation.

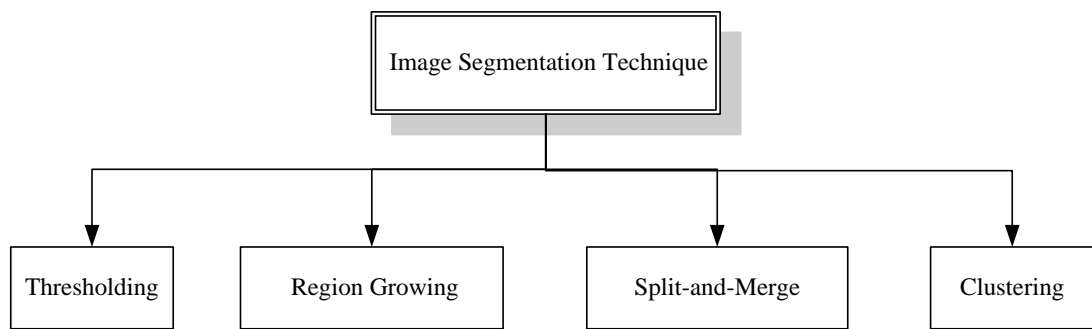
Generally, segmentation process is based on one of two basic properties of color, gray intensity, or texture, namely discontinuity around the edges, lines, and points in an image and similarity among the pixels in the same region (Gonzalez et al., 2004).



**Figure 2.1** Segmentation process for the (a) color, (b) gray intensity and (c) texture images.

### 2.2.2 Image Segmentation Techniques

Several techniques have appeared in the recent literature on image segmentation. In general, image segmentation techniques have been classified into four categories namely thresholding, region growing, split-and-merge and clustering as shown in Figure 2.2. The concept and some examples of application for threshold, region growing, split-and-merge, and clustering techniques are further dealt in Sections 2.2.2(a) to 2.2.2(d), respectively, wherein the basic concept, description on several clustering algorithms and their applications are discussed in detail.



**Figure 2.2** General classifications of the image segmentation technique.

#### 2.2.2(a) Thresholding Technique

In its basic form, thresholding techniques segment the scalar images by creating a binary partitioning of the image intensities. A thresholding procedure attempts to determine an intensity value, called threshold. The determination of threshold value is usually done based on the histogram of image intensity level. The pixels of an image are compared with the threshold value. It is marked as the pixels of the object (foreground) region if their value is greater than the threshold value or else they will



be marked as the background region. In mathematical form, the process can be described as:

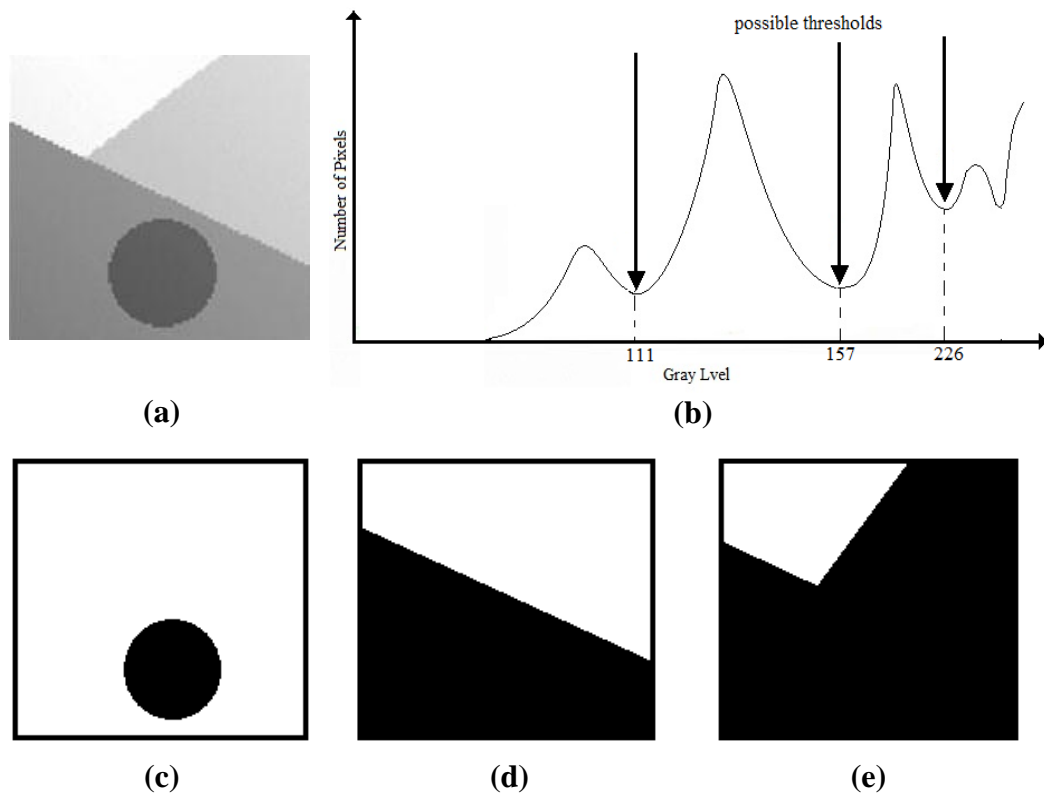
$$S(x, y) = \begin{cases} 1 & \text{if } p(x, y) > T \\ 0 & \text{if } p(x, y) \leq T \end{cases} \quad (2.1)$$

where,  $p(x, y)$  is a pixel of an image,  $S(x, y)$  is a marked or segmented pixel and  $T$  is the threshold value.

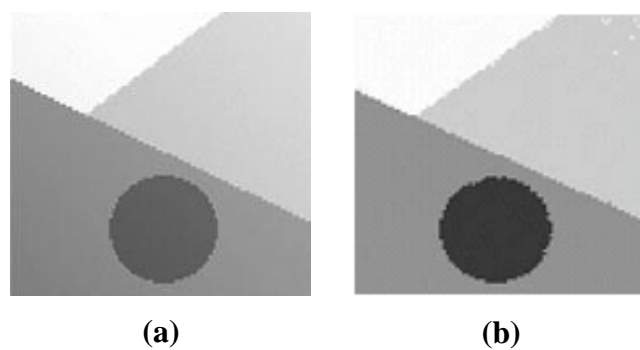
Usually, a value of 1 is given to object pixels while a value 0 is given to background pixels. Figure 2.3 shows an example of segmentation process of an image. Different threshold values will produce different segmented regions. Although, the image can be segmented into two regions by using only one threshold value, but it is possible that the image can be segmented into more than two regions by using multiple threshold values (Sahoo et al., 1988). As compared to single threshold technique, multiple threshold technique segments an image into more than one region of interest in a single phase of process, indirectly speeding up the process of segmentation (Hsu et al., 2010). Figure 2.4 illustrates the results obtained after applying thresholding technique with multiple threshold values.

As compare to other techniques, the thresholding is a simplest technique for obtaining a segmentation of an image with low computational time. However, the threshold does not take into account the spatial characteristics of an image. Thus, it becomes sensitive to noise, artifacts and inhomogeneity of an image. It only works

well and significantly separates the object from background if the histogram is bimodal (Tobias and Seara, 2002). One of its major drawbacks is that the solution is highly dependent on the initial threshold value.



**Figure 2.3** Implementation of the thresholding technique for a various threshold values (a) original image, (b) histogram of an image, (c) thresholded at 111, (d) thresholded at 157, and (e) thresholded at 226.



**Figure 2.4** Implementation of the thresholding technique with multiple threshold values (a) original image and (b) segmented image.

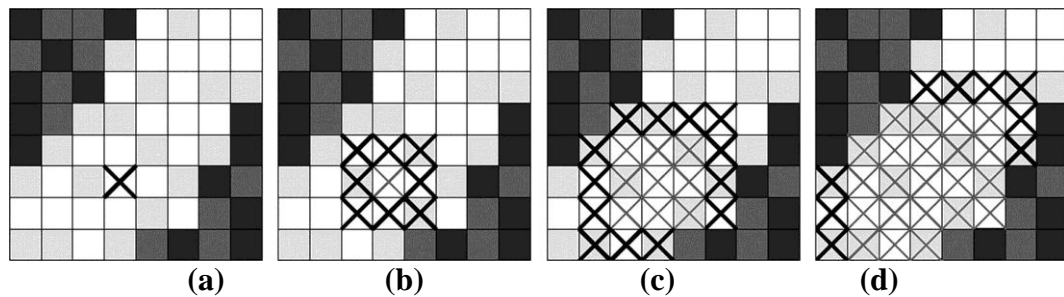
Due to its simplicity, the thresholding technique is often used in different applications as an initial image processing tool (Singleton and Pohost, 1997, Xu et al., 1999). The thresholding technique produces the binary image by means of which the region of interest can be easily recognizable. Therefore, it is widely used in medical imaging system for delineate the organs from collective tissues (Tobias and Seara, 2002). It is also used to accurately delineate the volume of object in the Positron Emission Tomography (PET) images (Drever et al., 2006, Drever et al., 2007). In remote sensing field, the thresholding scheme is applied to detect targeted object in a quick manner (Yaowen et al., 2010).

### **2.2.2(b)Region Growing Technique**

Region growing is a technique of extracting the region in an image that is connected according to some predefined criteria (i.e. gray intensity, color intensity or edges in the image). This basic approach begins from a predefined seed pixel in the predefined region seed that is to be segmented. Then, the region growing process includes all the neighboring pixels that are connected to predefined seed pixel based on predefined criteria. The process continues until all pixels have been considered or the produced region cannot be grown anymore. Figure 2.5 illustrates the implementation of region growing technique, while Figure 2.6 shows an example of region growing application for segmentation process.

The region growing technique offers several advantages such as simple implementation, low computational time and has capability of providing parallel

clustering process. However, it is sensitive to noise and variation of the intensities in an image, leading to occurrence of segmented regions with small holes or disconnected (Sun and Bhanu, 2009). In addition, the region growing technique requires manual interaction to obtain appropriate initial seed pixels, which is too subjective and time consuming especially for a complex image (i.e. medical images).



**Figure 2.5** Implementation of region growing technique (a) location of initial seed pixel, (b), (c), and (d) the growing process of seed pixel to form a region.



**Figure 2.6** The example of region growing application (a) original image (b) segmented image.

In term of applications, the region growing technique is often applied for medical images, where the interested region will be grown and distinguished from the surrounding region called background. Pohle and Toennies (2001) have successfully employed the region technique to segment out the kidney cortex, the liver and the passable lumen in an aortic aneurysm from CT scan images. In 2005, Mat Isa applied

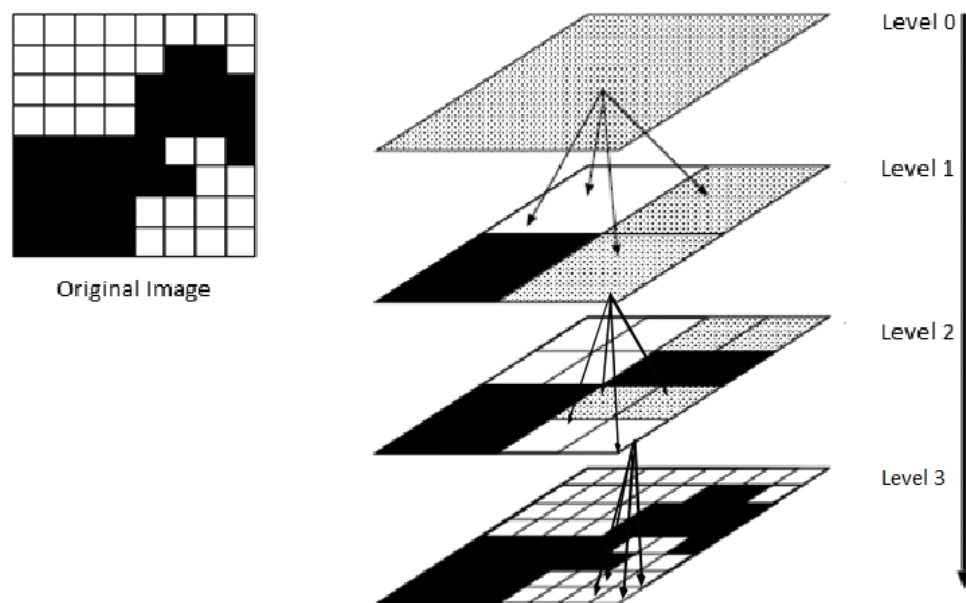
the region growing on Pap smear images, in order to segment the cytoplasm and nucleus of a cell. In addition, Mazonakis et al. (2001) proposed the region growing technique for the segmentation of prostate cancer on CT images. All of these applications could highlight certain regions on medical images for easier screening process by doctor. Other than that, the region growing process has also been applied in remote sensing (Abad et al., 2000, Dare, 2005, Kunte and Bhalchandra, 2010), satellite images (Bins et al., 1996, Derrien and Le Gléau, 2007, Anil and Natarajan, 2010) and face recognition (Jin et al., 2007).

### **2.2.2(c) Split-and-Merge Technique**

Split-and-Merge technique is a recursive partitioning technique, where each cluster (as a root of a tree) is firstly split along its co-ordinate axes to yield four sub-clusters. If the obtained sub-clusters are non-homogeneous (i.e. not equal to the predefined threshold value), then these sub-clusters are further split into four sub-squares until all pixels have been considered. On the other hand, if at least two out of these four sub-squares are homogenous, they can be merged to form a bigger cluster. This process continues recursively until no further splitting or merging process are possible (Pavlidis and Horowitz, 1974, Horowitz and Pavlidis, 1976). The abovementioned process could be illustrated as shown in Figure 2.7.

Similar to thresholding and region growing techniques, the split-and-merge technique is commonly applied as segmentation tool due to its simple implementation. Although its performance is comparable with others, it suffers

several limitations. The split-and-merge process is laborious and requires high computational time as the process must be applied to the whole image (Faruquzzaman et al., 2009). Higher processing time is required if bigger image size is under consideration. Other limitations are its segmentation performance subjectively depends on the settling of initial threshold value (Faruquzzaman et al., 2009), the final solution could probably converge to insignificant optimum location (Faruquzzaman et al., 2008) and inefficient to extract certain shapes of regions (specifically for the irregular shape) with smooth boundary and edges (Faruquzzaman et al., 2008).



**Figure 2.7** Implementation of the split-and-merge technique.

Based on the latter limitation, the split-and-merge technique is only applied for specific application (Faruquzzaman et al., 2009). One of the applications is

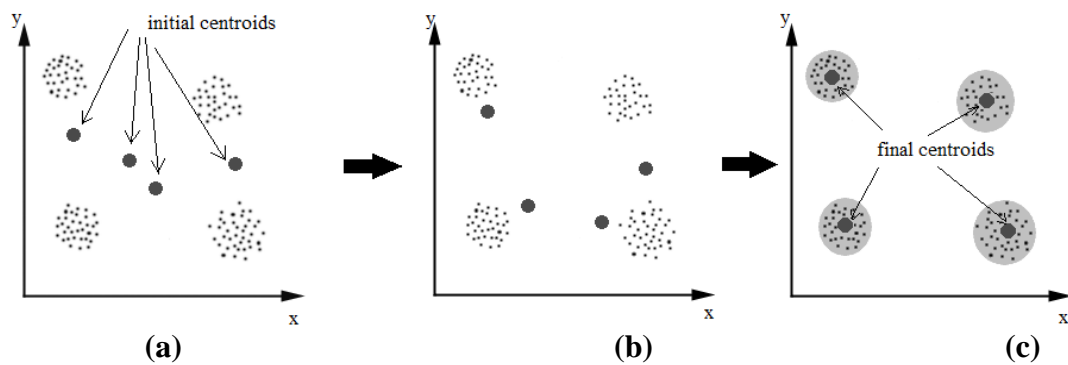
segmentation of the structure of building in the remote sensing images (Rau and Chen, 2003, Khoshelham et al., 2005). In that application, the structural information is essential for the building maintenance and urban planning. In addition, the split-and-merge technique is also applied in medical imaging (Manousakas et al., 1998, Nedzved et al., 2000), satellite imaging (Lucieer and Stein, 2002, Devereux et al., 2004), and face recognition (Sengupta et al., 2000, Meethongjan et al., 2010) areas.

#### **2.2.2(d)Clustering Technique**

Clustering is an unsupervised learning technique which classifies the pattern without accessing to any prior information about what it should be pursuing. For image segmentation, the clustering is employed for partitioning an image into disjoint clusters such that the pixels within the clusters are similar as possible and the pixels among the clusters are dissimilar as possible. In metric space, the similarity is defined by means of distance norm. The clustering segmentation process starts with randomly initialization of the centroids. Then, the centroid will be updated by iteratively computing a mean intensity for each cluster after assigning all pixels of an image to the nearest cluster. Figure 2.8 illustrates the implementation of the basic clustering process, while Figure 2.9 shows the resultant images after applying one clustering technique called FCM clustering.

Numerous researches have applied clustering technique as a segmentation tool. Those researches favored the clustering technique over other previously mentioned segmentation techniques due to its preferable characteristics such as, requires less

prior information of the properties of the data set and comparatively less effective to noise and artifacts. However, it could produce dead centroid because of the probability of centroids being trapped at non-active regions and its efficiency is depending on initial parameter (Mat Isa, 2005, Sulaiman and Mat Isa, 2010). The advantages of the clustering technique as compared to other abovementioned segmentation techniques are summarized in Table 2.1. Thus, this study will focus on development of new clustering algorithm for segmentation purpose. The detail explanation about the concept, advantages, disadvantages and application of several conventional clustering techniques will be discussed in Section 2.3.



**Figure 2.8** Illustration of basic clustering process; (a) initial condition when iteration=0, (b) condition of clustering process after iteration  $t$ , and (c) final condition after clustering process is completed.



(a)



(b)

**Figure 2.9** Result of segmentation process using FCM clustering technique; (a) original image (b) resultant image.

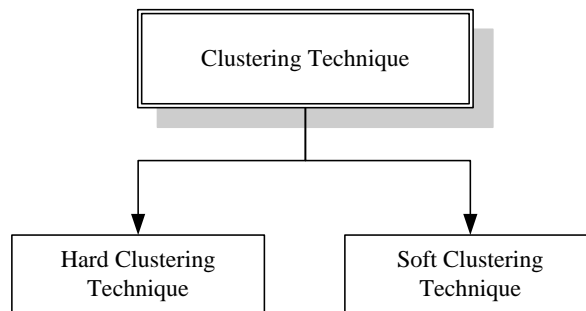


**Table 2.1** Characteristics comparison between segmentation techniques.

<b>Segmentation technique</b>	<b>Simplicity</b>	<b>Time complexity</b>	<b>Efficiency depend on initial parameters</b>	<b>Sensitive to noise and artifact</b>	<b>Require prior knowledge</b>	<b>Other major limitation</b>
Thresholding	Yes	Low	Yes	High	Yes	Only works well if the histogram is bimodal
Region Growing	Yes	Low	Yes	High	Yes	Intensity variation in an image causing over-segmentation
Split-and-Merge	Yes	High	Yes	High	Yes	Fail to segment the region with smooth boundary and edges
Clustering	Yes	Low	Yes	Low	No	Centroid being trapped at non-active region and produces empty cluster

## 2.3 Clustering-based Segmentation Technique

Clustering is a task for which different kinds of unsupervised learning algorithms have been introduced. Generally, the clustering technique is classified into two categories i.e. the hard clustering and soft clustering techniques as shown in Figure 2.10.



**Figure 2.10** General classification of clustering technique.

### 2.3.1 Hard Clustering Technique

The hard clustering technique employs the hard membership function during the clustering process. Due to this hard membership function, each pixel will be restricted to become a member of only one cluster. The K-Means (KM) algorithm is the best known clustering algorithm with hard membership function (MacQueen, 1967). The KM algorithm starts with initialization of cluster's center value called centroid. All pixels are assigned to the nearest cluster according to Euclidean distance and the centroid values of all clusters are then updated using specific statistical procedure. This iterative process is continued until no significant changes occur in the centroid value of the clusters. At the end of this iterative process, the solution should be able to converge to the global optimum solution.

The KM is not robust to the outliers and different initial parameters could yield different final solution (Bradley and Fayyad, 1998). Usually, the final solution is converged to poor local minimum location (Kaufman and Rousseeuw, 1990, Ester et al., 1995, Sulaiman and Mat Isa, 2010). Based on the work done by the Kaufman and Rousseeuw, (1990), the labeling of the cluster by the mean value of the cluster pixels is liable for outlier problem. This is because, in the KM clustering, the centroid is too sensitive to the far located pixels in a cluster, which could result in the closest cluster to become empty cluster or dead centroid. Thus, the final solution may converge to the poor local minima. In order to reduce the problem, the Partitioning Around Medoids (PAM) was proposed by the Kaufman and Rousseeuw, (1990). Based on PAM clustering algorithm, the labeling of cluster using medoid value is proven to be more effective than mean value to decrease the outlier sensitivity. A medoid is defined as a representative member of a cluster with a data set whose average dissimilarity to all the members in the cluster is minimal (Kaufman and Rousseeuw, 1990). However, the PAM is not robust to solve the initialization problem, which also encourages the empty cluster hitch. Furthermore, the computational complexity of PAM is very high as compared to KM and is not very effective in reducing outlier sensitivity for the large data (Barioni et al., 2008). Another approach called K-Medians clustering algorithm was employed for a better image segmentation (Anderson et al., 2006). The K-Medians algorithm is a variation of K-Means algorithm, where the median is employed to measure the centroid of all clusters. Formally, the objective of the K-Medians is to minimize the mean absolute deviation (MAD). The median or MAD is less sensitive to outlier as compare to the mean and medoid. However, the K-Median is also sensitive to initialization and may converge to poor local minimum location.

In the year 2000, a method called Moving K-Means (MKM) algorithm was proposed to resolve the convergence problem of the KM algorithm (Mashor, 2000). The research claimed that the dead centroid (i.e. empty cluster) and the trapped centroid at non-active region or between the two active centroids are the main reasons causing the solution of the KM to converge at poor local minima (Mashor, 2000). The empty clusters are obtained if the centroids have been initialized far away from the active regions or between two centroids of an active region. Due to both phenomena, the centroids may not be updated in the next iteration and may trap at non-active region. This unbalanced partitioning of the data (i.e. image) is overcome by using fitness criteria introduced in the MKM algorithm. According to this criterion, the fitness of each cluster is continuously verified during the updating process of cluster's centroid. In case of the predetermined unsatisfied condition, the cluster with the lowest fitness value gives up all its member pixels and moves towards the active region by getting the members of the cluster with the highest fitness value (i.e. which have less value than the cluster centroid). This approach keeps the variance of the entire clusters quite similar and becomes less effective to initialization process.

Although the MKM algorithm could significantly avoid the empty cluster, but it cannot reduce the trapped centroid problem as well as may converge to poor local minimum location. Mat Isa et al., (2010) pointed out that the improper transferring of member pixels of the cluster with the highest fitness to the cluster with the lowest fitness value is the main reason of inefficient performance of the MKM algorithm to converge to global optimum location. In order to rectify this problem, the Adaptive